

10/507111

1

Video Coding

5 This invention relates to video coding. In particular this invention relates to a method and apparatus for transmitting video data, and a method and system for compensating for transmission errors in a video data stream.

10 A video sequence consists of a series of still pictures or frames. Video compression methods are based on reducing the redundant and perceptually irrelevant parts of video sequences. The redundancy in video sequences can be categorised into spectral, spatial and temporal redundancy. Spectral redundancy refers to the similarity between the different colour components of the same picture. Spatial redundancy results from the similarity between neighbouring pixels in a picture. Temporal redundancy exists because objects appearing in the previous image are also likely to appear in the current image.

15 Compression can be achieved by taking advantage of this temporal redundancy and predicting the current picture from another picture, termed an anchor or reference picture. Further compression may be achieved by generating motion compensation data that describes the displacement between areas of the current picture and similar areas of the referenced picture.

20

Sufficient compression cannot usually be achieved by only reducing the inherent redundancy of the sequence. Thus, video encoders may also try to reduce the quality of those parts of the video sequence which are subjectively less important. In addition, the redundancy of the encoded bit-stream may be

25 reduced by means of efficient loss of coding of compression parameters and coefficients. The main technique is to use variable length codes.

Video compression methods typically differentiate between pictures that utilise temporal redundancy reduction and those that do not. Compressed pictures that do not utilise temporal redundancy reduction methods are usually called INTRA-frames, I-frames or I-pictures. Temporally predicted images are usually forwardly predicted from a picture occurring before the current picture and are called INTER or P-frames. In the INTER-frame case, the current picture is predicted from a reference picture, usually using a motion compensation technique, so generating prediction error data representing the differences between the two frames.

A compressed video clip typically consists of a sequence of pictures, which can be roughly categorised into temporally independent INTRA pictures and temporally differentially coded INTER pictures. As the compression efficiency in INTRA pictures is normally lower than INTER pictures, INTRA pictures are used sparingly, especially in low-rate applications.

A video sequence may consist of a number of scenes or shots. The picture contents may be remarkably different from one scene to another, and therefore the first picture of the scene is typically INTRA-coded. There are frequent scene changes in television and film material, whereas scene cuts are relatively rare in video-conferencing. In addition, INTRA pictures may typically be inserted periodically to stop temporal propagation of transmission errors in a reconstructed video signal and/or to provide random access points to a video bit-stream.

Compressed video is easily corrupted by transmission errors, mainly for two reasons. Firstly, due to the utilisation of temporal predicted differential

decoding (INTER frames), an error is propagated both spatially and temporally. In practice this means that, once an error occurs, it is easily visible to the human eye for a relatively long time. Especially susceptible are transmissions at low bit-rates where there are only a few INTRA-coded frames, so temporal error propagation is not stopped for some time. Secondly, the use of variable length codes increases the susceptibility to errors. When a bit error alters the code word, the decoder will lose code word synchronisation and also decode subsequent error-free code words (compressing several bits) incorrectly until the next synchronisation or start code. A synchronisation code is a bit pattern which cannot be generated from any legal combination of other code words and such codes are added to the bit-stream at intervals to enable re-synchronisation. In addition, errors occur when packets of data are lost during transmission which may produce visible errors in the image. For example, in video applications using the unreliable UDP transport protocol in IP networks, network elements may discard parts of the encoded video bit-stream.

There are many ways for the receiver to address the corruption introduced in the transmission path. In general, on receipt of a signal, transmission errors are first detected and then corrected or concealed by the receiver. Error correction refers to the process of recovering the erroneous data perfectly as if no errors had been introduced in the first place. Error concealment refers to the process of concealing the effect of transmission errors so that they are hardly visible in the reconstructed video sequence. Typically some amount of redundancy is added by the source or transport coding in order to help correct error detection, correction and concealment.

Current video coding standards define a syntax for a self-sufficient video bit-

stream, for example ITU-T recommendation H.263, "Video Coding for Low Bit-Rate Communication". This standard defines a hierarchy for bit-streams and correspondingly for image sequences and images.

- 5 In conventional systems, as shown in Figure 1, to reduce bandwidth when transmitting a video signal between two points, it is common to compress the picture frames by exploiting the spatial differences within a frame and the temporal differences between frames. The resulting signal is termed the "play stream". The video signal is transmitted as a series of packets of information.
- 10 The compression takes place in an encoder and the signal is then transmitted to a remote site where a decoder restores the image.

The loss or corruption of a packet of data will result in a mismatch between the encoder and decoder which typically appears as a visual error on the screen, for example, part of a moving object is "left behind". This error normally persists

15 until it is cleared or "cleaned" with a frame that is not predicted from a previous picture, such as an INTRA-frame.

In a system where a number of decoders are driven from one encoder, it is

20 usually beneficial to insert regular INTRA-frames in the play stream. However, this results in a loss of efficiency to all decoders as the INTRA-frames require a higher bit-rate than motion compensated frames.

Another common technique is to provide adequate buffering at the decoder so

25 that lost packets can be re-transmitted. However, this will produce delays at the decoder which may not be acceptable.

The present invention is directed to overcoming or substantially ameliorating the above problems.

5 According to a first aspect of the present invention, there is provided a method of transmitting video data, comprising the steps of:

generating a first video data stream:

generating a second video data stream comprising a plurality of frames each predicted from a reference frame;

transmitting data from the first stream to a receiver;

10 on receiving from the receiver an indication that data in the first stream is corrupted, transmitting data from the second stream to the receiver.

According to a second aspect of the present invention there is provided a method for compensating for transmission errors in a video data stream comprising:

15 transmitting a first video data stream from a transmitter to a receiver,
detecting corrupted data in the transmitted data stream,
generating an indication that data is corrupted, and
20 in response to the indication that the data is corrupted, transmitting data from a second video data stream predicted from a reference frame.

Preferably, the method further comprises reverting to the first video data stream after transmitting the data from the second video data stream.

25 In a preferred embodiment, the step of detecting corrupted data is carried out at the receiver, and preferably, the step of generating an indication that data is corrupted is carried out at the receiver.

Preferably, the step of generating an indication that data is corrupted includes the receiver generating an indication signal and transmitting the indication signal to the transmitter.

5

In a preferred embodiment, the step of transmitting data from the second video data stream is performed at the transmitter, the transmitted data from the second video data stream being received by the receiver.

10 According to a third aspect of the present invention, there is provided apparatus for transmitting video data, comprising:

an encoder for generating a first video data stream, the encoder further arranged for generating a second video data stream comprising a plurality of frames each predicted from a reference frame;

15 a transmitter for transmitting data from the first stream to a receiver;
means for receiving from the receiver an indication that data in the first stream is corrupted;

the transmitter upon receiving the indication is arranged for transmitting data from the second stream to the receiver.

20

Preferably, the transmitter is further arranged for reverting back to transmitting data from the first stream after data from the second stream has been transmitted to the receiver.

25 According to a fourth aspect of the present invention there is provided a system for compensating for transmission errors in a video data stream comprising:

a transmitter for transmitting a first video data stream,

a receiver for receiving the first video data stream,
means for detecting corrupted data in the first data stream, and
means for transmitting data from a second video data stream predicted
from a reference frame after detection of corrupted data in the first video data
5 stream.

Preferably, the means for detecting the corrupted data in the first video stream is
at the receiver, and, preferably, the transmitter is operable to transmit the data
from the second video data stream to the receiver after detection of corrupted
10 data in the first video data stream.

A preferred embodiment of the invention aims to provide a correction without
an increase in bandwidth by replacing lost packets of information with packets
from a fixed reference side stream, rather than inserting extra INTRA-frames.
15

The invention will now be described by way of example only with reference to
the accompanying drawings, in which:

Figure 1 is a block diagram showing the effect of packet loss in a conventional
20 video streaming system;

Figure 2 is a block diagram showing a conventional frame sequence with
INTRA (I) frames inserted;

25 Figure 3 is a block diagram showing a fixed reference side stream according to
an embodiment of the invention in which picture frames are predicted from a
single '0th' frame;

Figure 4 is a block diagram showing a frame sequence with feedback to overcome the effects of packet loss; and

- 5 Figure 5 is a block diagram showing an example of two data streams using the error correction system and method embodying the present invention.

The transmission of frames of video signals in a conventional system is shown in Figure 1. A transmitter 1 includes a buffer for storing frames 2 to be transmitted and an encoder 4 for encoding packets of data from the frames 2
10 stored in the buffer. It should be noted that the encoder 4 can run in either of two modes – “live” encoding for when video data comes from a live source, or “off-line” encoding when the encoder 4 may have operated on some archived content, possibly some time before the transmitter is running. In either case, no
15 feedback needs to be sent to the encoder 4. A receiver 5 includes a decoder 6 for decoding packets of information 8 received from the transmitter and producing video frames 10 from these packets of information 8.

Figure 2 shows a sequence of transmission frames 12 with INTRA frames 14
20 (also referred to as I-frames) inserted at intervals to clean the picture. This is a standard technique.

A fixed reference side stream as used in the invention is shown in Figure 3. Each frame (picture) 18 is derived from the same single frame 16, denoted as
25 the ‘0th’ frame. This reference frame 16 may be an INTRA frame, which is produced using known techniques.

A sequence of frames 20 showing the use of feedback to clean frames after packet loss is illustrated in Figure 4. The sequence of frames 20 comprises an INTRA frame 21 and a series of transmitted compressed frames 22, in at least one of which packet loss or corruption 23 has occurred. The loss or corruption of information is reported back to the transmitter which sends a correcting packet of data from the side stream 32 predicted from the reference frame 21 to produce a cleaned frame 24.

Figure 5 shows a transmitted play stream 30 comprising a series of frames and a corresponding fixed reference side stream 32. A number of frames 34 in the play stream 30 may contain missing or corrupted packets of information. If the receiver detects that a frame is corrupted, for example, when packet loss or corruption has occurred, this is signalled to the transmitter which then transmits packets from the fixed reference side stream 32 to clean the frame, and stop propagation of the errors.

The error compensation process according to a preferred embodiment of the invention will now be described by way of example.

A video signal to be transmitted is stored as a series of frames 2 in a buffer at the transmitter 1. The signal is encoded in the conventional manner by the encoder 4 and is transmitted as a series of packets of data 8 constituting a play stream 30 to one or more receivers 5. At the same time, the transmitter 1 produces a fixed reference side stream 32 in which the frames are all predicted from the same INTRA-frame rather than each being produced from the previous transmitted frame. At the receiver 5, the packets in the play stream 30 are decoded by the decoder 6 to recover the images.

If the receiver detects that a frame 34 is corrupted, for example, when packet loss or corruption has occurred, the receiver 5 sends a signal to the transmitter 1 notifying the transmitter of the error. The transmitter 1 then switches mode and, instead of sending the next packet from the play stream 30, the transmitter 1 sends a corresponding packet from the side stream 32. The packet from the side stream is predicted from a fixed reference frame instead of the preceding play stream frame. Thus, a cleaned frame 24 is produced at the receiver. This is shown in Figures 4 and 5. The system then reverts to the normal play stream 30.

Figure 5 shows how an entire frame's worth of information can be transmitted to produce a cleaned frame for use with subsequent play stream packets. However, the receiver or transmitter could also calculate which parts of the frame are missing or corrupted and transmission of the data from the side stream could be limited to the part of the frame (for example a GOB – group of blocks) necessary to clean the part of the frame containing errors rather than clean the whole frame.

The fixed reference side stream 32 illustrated in figures 3 and 5 differs from the play stream 30 in that all frames in the side stream are predicted from a single previous reference frame (picture), that is, the frames in the side stream 32 are obtained by comparing the current frame with the reference frame. This is in contrast to the play stream 30 in which the current frame is normally compared with the preceding frame in the play stream. The difference between the current frame and the reference frame will be transmitted when the transmitter transmits a frame from the side stream 32, as is the case when the receiver

notifies the transmitter of the detection an error. In a preferred embodiment, after receipt of data from the side stream 32, the receiver compares this with the stored reference (INTRA) frame to produce a cleaned current frame. The transmitter then reverts backs to transmitting data from the play stream 30.

5

The method and system for compensating for transmission errors embodying the invention is particularly advantageous as it does not require large amounts of buffering at the receiver and nor does it require a reduction in efficiency of the play stream to provide error resilience. The process embodying the invention permits decoding to continue once packet loss has occurred without significant delay in the play stream and whilst rebuffering occurs. This is particularly advantageous in low delay applications such as video conferencing applications in which any pause in transmission would be unacceptable. The method and system for compensating for transmission errors embodying the invention aims to provide quick recovery from loss or corruption and to minimise loss in quality which would result if conventional I-frames were used. Furthermore, the quality of the play stream is not compromised to provide extra resilience.

10

15

20

25

The invention is not intended to be limited to the video coding protocol or compression schemes mentioned above and in the drawings which are intended to be merely exemplary. The invention is applicable to any video coding protocol using temporal prediction, such as MPEG4 and H.263. Furthermore, whilst the invention has been described as being applicable to compensate for errors due to packet loss, it may also be applied to compensate for bit errors.